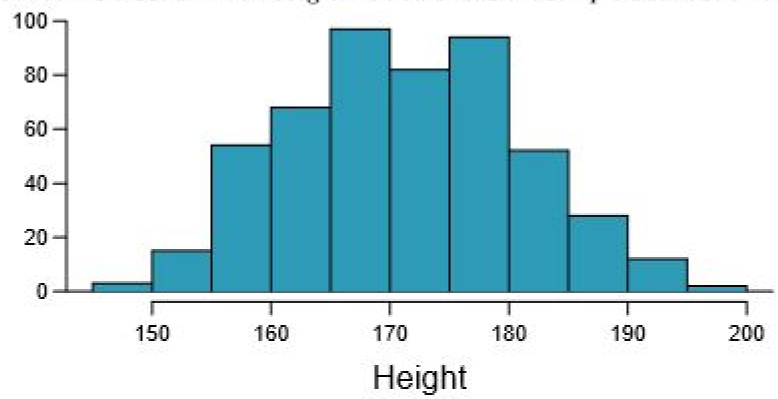
week_4_homework

Christophe

September 28, 2015

4.4 Heights of Adults

4.4 Heights of adults. Researchers studying anthropometry collected body girth measurements and skeletal diameter measurements, as well as age, weight, height and gender, for 507 physically active individuals. The histogram below shows the sample distribution of heights in centimeters.[38]



Min	147.2
Q1	163.8
Median	170.3
Mean	171.1
SD	9.4
Q3	177.8
Max	198.1

(a) What is the point estimate for the average height of active individuals? What about the median?

The point estimate for the average height of active individuals is `171.1` and the median is `170.3`.

(b) What is the point estimate for the standard deviation of the heights of active individuals? What about the IQR?

The point estimate for the standard deviation of hieghts of active individuals is `9.4` and the IQR is 14.

(c) Is a person who is 1m 80cm (180 cm) tall considered unusually tall? And is a person who is 1m 55cm (155cm) considered unusually short? Explain your reasoning.

A person who is 180 cm would have a Z Score of 0.95 and a person who is 155 cm would have a Z score of -1.71. Since both are at not more than 2 standard deviations from the mean then I would not consider them exceptionally tall or short, I would say that the shorter person is approaching the limit but not quite there.

(d) The researchers take another random sample of physically active individuals. Would you expect the mean and the standard deviation of this new sample to be the ones given above? Explain your reasoning.

I would not expect the new sample to have the same values as given in this example. The researchers are sampling a random set of physically active individuals and the results will vary with each group, because it would be impossible to sample the entire population. However, I would expect another sample of people to be relatively close to the values provided from this example.

(e) The sample means obtained are point estimates for the mean height of all active individuals, if the sample of individuals is equivalent to a simple random sample. What measure do we use to quantify the variability of such an estimate (Hint: recall that

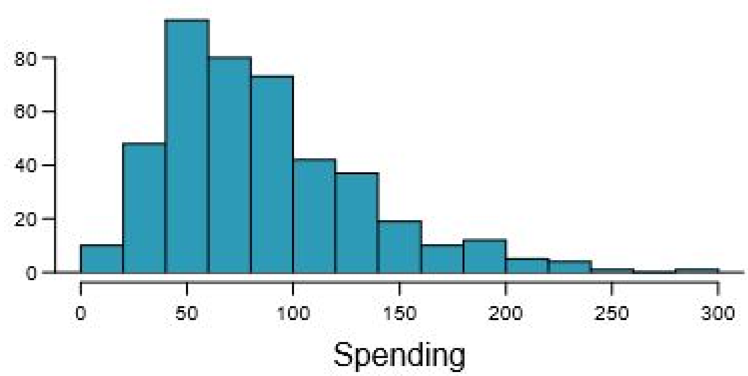
$$SD\bar{x} = \frac{\sigma}{\sqrt{n}}$$

)? Compute this quantity using the data from the original sample under the condition that the data are a simple random sample.

We use the standard error which is 0.417 for this sample's mean.

4.14 Thanksgiving Spending: Part I

4.14 Thanksgiving spending, Part I. The 2009 holiday retail season, which kicked off on November 27, 2009 (the day after Thanksgiving), had been marked by somewhat lower self-reported consumer spending than was seen during the comparable period in 2008. To get an estimate of consumer spending, 436 randomly sampled American adults were surveyed. Daily consumer spending for the six-day period after Thanksgiving, spanning the Black Friday weekend and Cyber Monday, averaged \$84.71. A 95% confidence interval based on this sample is (\$80.31, \$89.11). Determine whether the following statements are true or false, and explain your reasoning.



(a) We are 95% confident that the average spending of these 436 American adults is between \$80.31 and \$89.11.

False, we are 100% confident that the average spending of this sample of 436 American adults is between \$80.31 and \$89.11.

(b) This confidence interval is not valid since the distribution of spending in the sample is right skewed.

False, since the data is not strongly skewed and the number of observations is sufficiently large we can use normal methods.

(c) 95% of random samples have a sample mean between \$80.31 and \$89.11.

False, 95% confidence is not a probability that samples will fall in the mean. We could take random samples and find that in practice 50% fall within the interval, however, from our original sample we were 95% confident that random sample means would fall in the interval.

(d) We are 95% confident that the average spending of all American adults is between \$80.31 and \$89.11.

True, the confidence interval allows us to say that with 95% confidence the true mean should fall within the interval parameters.

(e) A 90% confidence interval would be narrower than the 95% confidence interval since we don't need to be as sure about our estimate.

True, a lower confidence would allow us to have a more narrow range but we have to sacrifice accuracy for the more narrow range.

(f) In order to decrease the margin of error of a 95% confidence interval to a third of what it is now, we would need to use a sample 3 times larger.

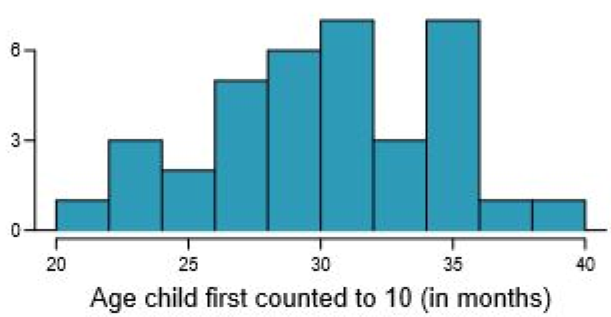
False, we use the square root of the number of observations to determine the standard error. In order to decrease by a third we would need to same 3ˆ2 = 9 times the initial sample.

(g) The margin of error is 4.4

TRUE, the standard error is half the width of the confidence interval at 4.4.

4.24 Gifted children, Part I.

4.24 Gifted children, Part I. Researchers investigating characteristics of gifted children collected data from schools in a large city on a random sample of thirty-six children who were identified as gifted children soon after they reached the age of four. The following histogram shows the distribution of the ages (in months) at which these children first counted to 10 successfully. Also provided are some sample statistics.[43]



n	36
min	21
mean	30.69
sd	4.31
max	39

(a) Are conditions for inference satsified?

The sample size is greater than 30 but less than 10% of the total population, the sample appears to be normally distributed. Assuming all else is sufficient we could conclude that the conditions for inference are satsified.

(b) Suppose you read online that children first count to 10 successfully when they are 32 months old, on average. Perform a hypothesis test to evaluate if these data provide convincing evidence that the average age at which gifted children first count to 10 successfully is less than the general average of 32 months. Use a significance level of 0.10.

$$H_o : \mu = 32 months$$
$$H_A : \mu > 32 months$$

```
gc_mean <- 30.69
gc_sd <- 4.31
nongc_mean <- 32

z_score <- ((gc_mean - nongc_mean)/gc_sd)

p_value4.24 <- (1 - pnorm(z_score))

paste("The p value is", round(p_value4.24,4),
      "which is greater than the",
      " significance value of .10 so we fail to reject",
      " the null hypothesis")
```

[1] "The p value is 0.6194 which is greater than the significance value of .10 so we fail to reject the null hypothesis"

(c) Interpret the p-value in context of the hypothesis test and the data.

The p-value is the probability of observing a sample mean of `30.69` for a sample of 36 children is 61.9%. That is, if the null hypothesis is true, we would often see a mean of this size.

(d) Calculate a 90% confidence interval for the average age at which gifted children first count to 10 successfully.

```
mean_gfmocount <- 30.69
sd_gfmocount <- 4.31
n <- 36
con_level_multipler <- 1.65

paste("The confidence interval for average",
      " months before a gifted child counts is (",
      round( mean_gfmocount - ((sd_gfmocount/sqrt(n))
                               * con_level_multipler),2) ,
      ", ", round( mean_gfmocount + ((sd_gfmocount/sqrt(n))
                                     * con_level_multipler), 2) ,
      ")", sep = "")
```

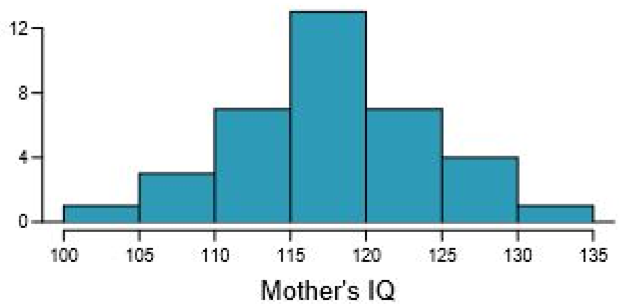
[1] "The confidence interval for average months before a gifted child counts is (29.5, 31.88)"

(e) Do your results from the hypothesis test and the confidence interval agree? Explain.

No, they seem to disagree. The p-value clearly shows that the finds of the study do not provide statistically significant findings. That is, we would expect to observe the same mean found with gifted children as with the population as a whole quite often. Whereas, the confidence interval provides a range less than the general population mean which could be misleading. The two provide conflicting opinions of significance.

4.26 Gifted children, Part II.

4.26 Gifted children, Part II. Exercise 4.24 describes a study on gifted children. In this study, along with variables on the children, the researchers also collected data on the mother's and father's IQ of the 36 randomly sampled gifted children. The histogram below shows the distribution of mother's IQ. Also provided are some sample statistics.



n	36
min	101
mean	118.2
sd	6.5
max	131

(a) Perform a hypothesis test to evaluate if these data provide convincing evidence that the average IQ of mothers of gifted children is different than the average IQ for the population at large, which is 100. Use a significance level of 0.10.

$$H_o : \mu = 100IQ$$

$$H_A : \mu > 100IQ$$

```
mean_normalIQ <- 100
mean_gfIQ <- 118.2
sd_giftedIQ <- 6.5

z_score <- ((mean_gfIQ - mean_normalIQ)/ sd_giftedIQ)

p_value4.26 <- (1 - pnorm(z_score))

paste("The p value is", round(p_value4.26,4),"which is less than",
      " the significance value of .10 so we reject the null hypothesis")
```

[1] "The p value is 0.0026 which is less than the significance value of .10 so we reject the null hypothesis"

(b) Calculate a 90% confidence interval for the average IQ of mothers of gifted children.

```r
mean_IQ <- 118.2
sd <- 6.5
n <- 36
con_level_multipler <- 1.65

paste("The confidence interval for average IQ of ",
      " mothers of gifted children is (",
      round( mean_IQ - ((sd/sqrt(n)) * con_level_multipler),2) ,
      ", ", round( mean_IQ + ((sd/sqrt(n)) *
                              con_level_multipler), 2) ,
      ")", sep = "")
```

[1] "The confidence interval for average IQ of mothers of gifted children is (116.41, 119.99)"

(c) Do your results from the hypothesis test and the confidence interval agree? Explain.

> The results do agree, the hypothesis test provides evidence that mothers of gifted children will themselves have high IQs. Also, the confidence interval provides a interval that is greater than the average IQ. The two measures provide convincing evidence to reject that all gifted children have an equal probability of having a mother of average intelligence.

4.34 CLT.

Define the term "sampling distribution" of the mean, and describe how the shape, center, and spread of the sampling distribution of the mean change as sample size increases

> The term sampling distribution represents the distribution of the point estimates based on samples of a fixed size from a certain population. The shape, center, and spread of the sampling distribution becomes more normal as the sample size increases. Essentially, this means that the population estimates become more accurate with repeated sampling.

4.40 CFLBs.

A manufacturer of compact fluorescent light bulbs advertises that the distribution of the lifespans of these light bulbs is nearly normal with a mean of 9,000 hours and a standard deviation of 1,000 hours.

(a) What is the probability that a randomly chosen light bulb lasts more than 10,500 hours?

> The probability that a randomly chosen light bulb will last more than 10,500 hours is 6.67%

(b) Describe the distribution of the mean lifespan of 15 light bulbs.

> The distribution should be near normal and symetrical. The distribution would be N(9000, 258.199) However, 15 lightbulbs is a very low sample size so we could see some large done flictions.

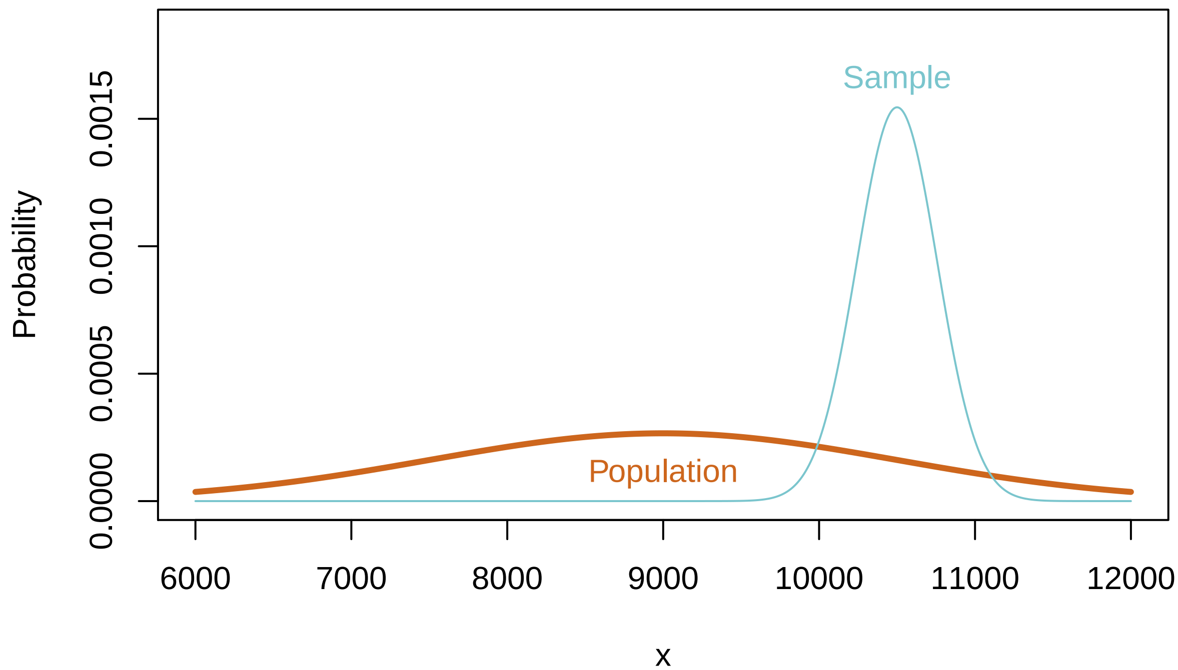
(c) What is the probability that the mean lifespan of 15 randomly chosen light bulbs is more than 10,500 hours?

```r
mean_normal <- 9000
mean_sample <- 10500
sd_sample <- 1000/(sqrt(15))

z_score <- ((mean_sample - mean_normal)/ sd_sample)

paste("The probability that 15 randomly sampled light",
      " bulbs would have a mean at 10,500 hours or more is",
      percent(1 - pnorm(z_score)) )
```

[1] "The probability that 15 randomly sampled light bulbs would have a mean at 10,500 hours or more is 0.000000313%"

(d) Sketch the two distributions (population and sampling) on the same scale.

```r
library(ggplot2)
x  <- 6000:12000
a <- dnorm(x, mean = 9000, sd = 1500)
b <- dnorm(x, mean = mean_sample, sd = sd_sample)
plot(x,a,type="l",col="chocolate3", lwd = 3, ylim=c(0,1.2*max(a,b)), ylab="Probability")
text(9000, b[which(x==9000)], "Population", col="chocolate3", pos=3)
lines(x,b,type="l", col="cadetblue3")
text(mean_sample, b[which(x==mean_sample)], "Sample", col="cadetblue3", pos=3)
```



(e) Could you estimate the probabilities from parts (a) and (c) if the lifespans of light bulbs had a skewed distribution?

> I could not estimate the probabilities from part (a) if it had a skewed distribution because the techinique for determining the probability relies on the distribution being near normal. Also, I could not estimate (c) since the sample size is so small that a skew distribution would be impossible to estimate if the distribution wasn't already near normal.

4.48 Same observation, different sample size.

Suppose you conduct a hypothesis test based on a sample where the sample size is n = 50, and arrive at a p-value of 0.08. You then refer back to your notes and discover that you made a careless mistake, the sample size should have been n = 500. Will your p-value increase, decrease, or stay the same? Explain.

> Typically, the p value would decrease as the sample size begins to increase because we would have stronger evidence to reject the null hypothesis. However, we could expect an increase in p value if the null hypothesis is true.